

2019/12/10 帝京大学・腫瘍内科学特論

臨床試験論文の読み方

北海道大学 医学統計学
横田 勲

1

講義内容

- ▶ 出版バイアスと臨床試験登録
- ▶ CONSORTガイドライン
 - ▶ 統計的仮説検定とサンプルサイズ
 - ▶ 解析対象集団

2

有意でないと論文にならない

JCO 2008; 2013-9.

3

有意でないと論文にならない？

Lancet Oncol 2012; 1225-33.

4

ネガティブな論文は出づらい

- ▶ 企業にとっては出したくない
 - ▶ 販売促進に使づらい
 - ▶ メディカルライティング等費用もかかる
- ▶ 医師にとって
 - ▶ インパクトファクターの高い雑誌に出せない
 - ▶ だったら忙しいし、作成しない
- ▶ 世の中にとっては・・・

5

出版バイアス

- ▶ ポジティブな結果ばかり出版されるため全体では過大評価してしまう
- ▶ Funnel plot

各研究結果

サンプルサイズ (精度)

1000

100

10

0.4 0.7 1

ハザード比 (治療効果)

ネガティブな方向の研究結果が少ない

6

公正な出版のために

7

- ▶ 試験結果は必ず論文化
 - ▶ ヘルシンキ宣言2000,2002
 - ▶ 人を対象とする医学系研究に関する倫理研究
- ▶ そもそもどんな試験を実施中かを公開
 - ▶ 臨床試験登録(ICMJE; 2005)
- ▶ 誰がスポンサーで誰が執筆したか
 - ▶ COIの開示(ICMJE 統一投稿規定; 2008)
 - ▶ GPP3 (2015) 企業がスポンサー or 支援する場合
- ▶ 論文に必要な項目は何か
 - ▶ CONSORTガイドライン(1996-, 2010最新改訂)

ICMJE: 医学雑誌編集者国際委員会

7

臨床試験登録

8

- ▶ 科学的、倫理的、説明責任のため
臨床試験の概要に関する項目を開示
- ▶ WHO・ICTRPの一次レジストリ、
もしくは ClinicalTrials.govへの登録
 - ▶ 複数システムへの登録もOK
 - ▶ 日本はUMIN、JAPIC、JMACCT、jRCT
 - ▶ 国立保健医療科学院のポータルサイトを参照
- ▶ 発行される登録番号なしでは、
承認資料に使えない、論文化できない

8

The EQUATOR network

<http://www.equator-network.org/>

9

9

CONSORT statement

10

- ▶ ICMJEの活動成果
- ▶ 1996年初版
 - ▶ 2001年、2010年に改訂
- ▶ 群間並行ランダム化試験(RCT)を対象
 - ▶ クラスタランダム化試験や非劣性試験には
CONSORT拡張版
- ▶ 400以上の雑誌が支持
- ▶ 25項目のチェックリストと
フローダイアグラム

10

CONSORTの目的・方針

11

- ▶ 完全、明快かつ透明性の高い報告を発展
 - ▶ 著者とpeer reviewerによる査読
 - ▶ 読者による文献の批判的評価 を助ける
- ▶ 厳格な形式を守らせるものではない
 - ▶ 論文の構造の標準化を意図するものではない
 - ▶ 雑誌のスタイル、投稿規定、研究分野の慣習、
著者の好みなどに論文の形式はしたがうべき
- ▶ チェックリストは、試験の質スコアを
構築するために使うべきではない

11

CONSORT diagram

12

12

CONSORT チェックリスト

13

13

タイトル・抄録・イントロ

14

- 1a タイトルにランダム化比較試験であることを記載
- 1b 構造化抄録
 - ▶ 背景、試験デザイン、方法、結果、結論
- 2a 科学的背景と論拠
- 2b 目的や仮説

14

試験デザイン

15

3a 試験デザイン

- ▶ 並行群間か、要因実験か
- ▶ ランダム化の割付比も書く
 - ▶ ランダム化の詳細は8a,8bにて

3b 試験開始後の方法の重要な変更

- ▶ 理由を明らかにする

15

参加者

16

4a 適格規準・条件

- ▶ 選択基準は、治療を適用したい標的集団を表した規準
 - ▶ 試験であれば適応に関わる
- ▶ 除外基準は、評価にバイアスを招く被験者を除くための規準
 - ▶ 安全性が保証できないような条件をもつ人は除く

4b データ収集のセッティングと場所

- ▶ 時期や研究を行った国、地域、施設

16

介入

17

5 各群の介入

- ▶ 他の研究者、臨床家が再現可能となるよう詳細に述べる
 - ▶ 研究のオリジナリティーがあることも
- ▶ 介入スケジュールや手順まで述べる
 - ▶ × 試験薬剤を投与した
 - ▶ ○ 試験薬剤を連日投与する4週と、その後2週間の休薬期間を1サイクルとし、合計8サイクル行った

17

アウトカム①

18

6a 主要・副次評価項目

- ▶ 主要評価項目
 - ▶ その研究で最も明らかにしたい仮説を評価するための項目
 - ▶ 通常、type-Iエラーを厳しく制御する
- ▶ 副次評価項目
 - ▶ 主要ほどではないが、確認しておきたい評価項目
 - ▶ 主要で代替エンドポイントを用いたならば、真のエンドポイントを副次に含めることも
 - ▶ 「探索的」の名のもと、type-Iエラーの制御は行わないことが多い

18

アウトカム②

19

6b 評価項目の変更

- ▶ 診断基準の変更
- ▶ 「イベント」の定義の変更
 - ▶ 長い試験期間の中で変更することがある
 - ▶ がん分野ではあまり見かけない
- ▶ 主要評価項目の変更の場合
 - ▶ 特に慎重にせねばならない
 - ▶ そのための仮説を正しく示すことができるのか？
 - ▶ スポンサー側に都合のいい変更でないか？

19

20

2005.10 登録開始

2007.7 選択基準変更

- ▶ 463例登録済み
- ▶ 2008.10 主要評価項目をOSからPFSに
- ▶ 1299例登録済み
- ▶ 2009.6 1873例登録完了

▶ OSをみるための増悪後の盲検維持は infeasible

▶ 後治療の影響を考慮し、PFSをプライマリに

NEJM 2011; 2473-83

20

サンプルサイズ①

21

7a サンプルサイズの設定根拠

- ▶ 主要評価項目の検定に対応させてサンプルサイズ設計を行う

21

サンプルサイズの決め方

22

- ▶ 「群間差がない」場合に、検定で有意差がつく割合を有意水準以下に
- ▶ 「群間差がある」場合に、検定で有意差がつく割合を一定水準以上に
 - ▶ この割合を検出力という
 - ▶ 「群間差」には臨床的に意味のあるを設定
- ▶ 以上を満たす必要サンプルサイズを計算

22

統計的仮説検定

23

- ▶ 帰無仮説：「比較群間で差がない」
 - ▶ OSに関し、対照群に対する試験群のハザード比が1
- ▶ 帰無仮説が正しい場合、観測された結果がどの程度極端なものであるかを評価
 - ▶ ありうるパターンを全通り考えたうち、何番目に極端なものかを計算
 - ▶ これがP値

23

割付を並べ替え

24

観測された結果

| ID | OS time | 割付 |
|----|---------|----|
| 1 | 12 | 試験 |
| 2 | 7 | 対照 |
| 3 | 13 | 試験 |
| 4 | 3 | 対照 |
| 5 | 28* | 試験 |

ハザード比：0.5

割付をランダムに並べ替えれば、試験と対照が同じ治療効果の場合に観察されるハザード比の分布が判明

| ID | OS time | 割付 | 試験 | 対照 |
|----|---------|----|----|----|
| 1 | 12 | 試験 | 試験 | 対照 |
| 2 | 7 | 試験 | 対照 | 対照 |
| 3 | 13 | 試験 | 試験 | 試験 |
| 4 | 3 | 対照 | 対照 | 対照 |
| 5 | 28* | 対照 | 試験 | 試験 |

ハザード比：1.2

ハザード比：1.5

ハザード比：0.7

*打ち切り

24

P値

割付パターンを全通り並べ替え、推定したハザード比の分布

並べ替え分布全体のうち、観察されたハザード比は何%目にあるか

- ▶ パターン数が膨大なので、確率で表す
- ▶ 事前に定めた有意水準より小さければ、帰無仮説は間違いとする（棄却する）

25

P値と有意水準

- ▶ 有意水準は片側2.5%がよく用いられる
- ▶ P値が有意水準より小さい場合
 - 群間差があった
- ▶ P値が有意水準より大きい場合
 - ✗ 群間差がなかった
 - 群間差があると証明できなかった
 - 群間差があるかないか、分からない

26

αエラーとβエラー

- ▶ αエラー：帰無仮説は正しいのに棄却してしまう
- ▶ βエラー：帰無仮説は正しくないのに棄却しない

| | | |
|---------------|------------------------|------------------|
| | 本当は帰無仮説は・・・ | |
| | 正しくない | 正しい |
| 有意（棄却する） | 1-β 検出力power | α 有意水準 |
| 有意でない（棄却されない） | β | 1-α |

27

αエラーとβエラー

有意でない ← 有意

帰無仮説の下での治療効果 いき値 対立仮説の下での治療効果

βエラー αエラー

帰無仮説は正しくないのに、有意とならない確率 (=β) 帰無仮説は正しい仮定のもと、有意となる確率 (=α/2)

28

αエラーとβエラー

もしサンプルサイズを大きくして分散を小さくすると・・・

βエラーは小さくなる

29

適切なサンプルサイズ

- ▶ サンプルサイズをただ大きくすればよいのか？
 - ▶ 実際の研究では、多くの人数を集めることは時間的、金銭的に困難
 - ▶ わずかな差でも有意となってしまう、そのような差に臨床的意味があるのか？
- ▶ サンプルサイズが小さいとどうなるか？
 - ▶ 臨床的に意味のある差があるのに、有意であるという結果が得られにくくなる
 - ▶ 検出力不足

30

生存時間アウトカムに対する サンプルサイズ設計の必要項目

31

- ▶ 検定手法 たいていlog-rank検定
- ▶ 有意水準 たいてい片側2.5%
- ▶ 検出力 たいてい80%
- ▶ 以下の項目のうち2つ
 - ▶ ハザード比（期待する治療効果）
 - ▶ 対照群のハザード・〇年生存割合
 - ▶ 試験群のハザード・〇年生存割合
- ▶ 試験期間と患者登録期間
 - ▶ 生存時間アウトカム特有の項目

31

2群比較の優越性試験の場合

32

- ▶ 代表的な場合を表示

| アウトカム | 期待する治療効果 | 付加的情報 |
|-------|--------------------|--|
| 連続量 | 平均値の差 | ・標準偏差 |
| 二値 | リスク差、リスク比、 オッズ比 | ・いずれかの群の 発生割合 |
| 生存時間 | ハザード比 | ・いずれかの群の ハザードや生存率 ・試験期間 ・患者登録期間 |

32

サンプルサイズ設計の実際

33

- ▶ たいていは専用ソフトを用いる
 - ▶ SAS、JMP、nQuery、PASS、、、（有料）
 - ▶ SWOGのホームページ（無料）
- ▶ 複雑な仮説の場合はシミュレーション
 - ▶ 群間差がない下で、n人分のデータを発生し、検定を行い、有意な割合が有意水準以下
 - ▶ 期待する効果の下で、n人分のデータを発生、検定を行い、有意な割合が検出力以上
 - ▶ このnをいろいろ変えて検討

33

サンプルサイズ②

34

7b 中間解析と中止基準

- ▶ 有効性の中間解析
 - ▶ 治療法が想定以上に有効であった場合、早期に試験を中止して、早く世の中に出したい
- ▶ 無効性の中間解析
 - ▶ 治療法が想定よりはるかに効かなそうである場合、早期に試験を中止したい
- ▶ 安全性の中間解析（モニタリング）
 - ▶ 新薬開発ではほぼ必ず
 - ▶ この節でいう「中間解析」には通常含まれない

34

検定の多重性

35

- ▶ 中間解析で5%水準で検定
- ▶ 最終解析でも5%水準で検定
 - ▶ 治療が全く有効でない場合でも、有意差ありと結論してしまう機会が増える
 - ▶ Type-I エラーの増加

35

Bonferroni流の調整

36

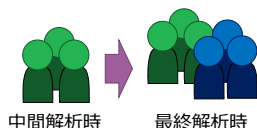
- ▶ 有意水準を検定の回数で割る
 - ▶ 2回検定をするなら、有意水準を2.5%に
 - ▶ 全体でtype-I エラーを5%水準に担保できる
- ▶ かなり効率を失う
 - ▶ 本当は差がある場合に、きちんと有意差ありと判断する機会を失う
 - ▶ 2回の検定は独立でないため
 - ▶ 最終解析時データの一部は中間解析時データ

36

グループ逐次法 group sequential method

37

- ▶ 中間解析からデータが追加されて、次回解析に用いられる特徴を考慮
- ▶ Lan-DeMets の α 消費関数法
 - ▶ O'Brien-Fleming型
 - ▶ Pocock型
- ▶ Haybittle-Peto法



37

ランダム割付

38

- 8a 割付の方法
- 8b 割付のタイプ
 - ▶ 層別割付？層別因子は？
 - ▶ 置換ブロック法？ブロックサイズは？
 - ▶ 最小化法？
- 9 割付表の管理
- 10 割付の実施者

38

盲検化

39

- 11a 誰が（何が）盲検されたか
 - ▶ 単盲検？二重盲検？
 - ▶ アウトカム評価を盲検？
 - ▶ PROBE(Prospective, Randomized, Open, Blinded-Endpoint)法
- 11b 盲検した内容はバレなかったか
 - ▶ プラセボは実薬とそっくりだったか
 - ▶ 試験群と対照群で評価は同等であったか
 - ▶ PD判定は施設であったか中央であったか

39

統計解析

40

- 12a 主要/副次評価項目の解析方法
 - ▶ 解析を再現できる記載
 - ▶ 統計解析計画書(SAP)は作った？
 - ▶ 信頼区間も表示すべき
- 12b サブグループ解析や調整解析
 - ▶ データを見る前に決めたか(pre-specified)、データを見た後に行ったか(post-hoc)
 - ▶ 割付因子は解析でも調整

40

対象者の流れ

41

- 13a 対象者の流れ
 - ▶ CONSORTダイアグラムをかく
- 13b 対象者の追跡不能と除外の理由
 - ▶ 群間のバランスが崩れると、解析結果にバイアスが生じる

41

対象者の募集と追跡期間

42

- 14a 参加者の登録期間と追跡期間
- 14b 試験終了/中止理由
 - ▶ 中間解析で有効中止した場合
 - ▶ 治療効果には潜在的にバイアスが入っている
 - ▶ もう少し追跡すると、治療効果が小さくなる場合がほとんど（平均への回帰）
 - ▶ バイアス補正法も提案されているが(*)、あまり使われない
 - ▶ たいてい、中間解析の有意水準と95%信頼区間を表示

* Tsiatis, Rosner and Mehta (1984), Emerson and Fleming (1990)

42

対象者のベースライン特性 43

15 人口統計学的、臨床的な特性

- ▶ ランダム化した場合、群間比較のために検定することは無意味
 - ▶ 偶然によって起こること
- ▶ 予後因子のバランスが崩れた場合、その影響の強さとバランスの度合いを考慮

43

解析対象集団 44

16 解析された人数

- ▶ 各群での実際に解析された人数を明らかに
 - ▶ 分子分母ともに
 - ▶ 何らかの理由により、アウトカムが観察されなかった場合があるため

44

FAS ? PPS ? SAS ? ① 45

- ▶ Full Analysis Set (最大の解析対象集団)
 - ▶ 基本的に登録された対象者全員
 - ▶ 割付通りに解析する (ITTの原則)
 - ▶ 全データ使用不可の同意撤回や、追跡データなし、登録後不適格判明を除外する定義も
- ▶ Per Protocol Set (プロトコル遵守集団)
 - ▶ プロトコル治療を行った対象者
 - ▶ 試験によって定義はまちまち
 - ▶ 実際の投薬別に解析してもよい

45

FAS ? PPS ? SAS ? ② 46

▶ Safety Analysis Set (安全性解析集団)

- ▶ プロトコル治療が行われた対象者全員
 - ▶ 基本的に登録後不適格判明も含める

46

Intention To Treat (ITT) の原則 47

- ▶ 治療方針による効果は、実際に受けた治療ではなく、**被験者を治療しようとした意図**に基づくことにより最もよく評価できる
ICH-E9 (統計)ガイドライン
- ▶ 2群比較の優越性試験では、保守的 (群間差がつかない) 方向のみバイアスが入った結果が得られる
 - ▶ 規制当局からすれば望ましい方向のバイアス

47

ITT原則の特徴 48

測定できない要因

ランダム割付 → 実際の治療 → 結果

- ▶ 割付後に、対象者の状態をみて、試験治療を行わなかった場合
 - ▶ 実際の治療内容で解析すると、治療効果を過大評価するおそれがある
 - ▶ 対照治療を受けても、割付通りの解析では、試験群として解析されるため、効果を過小評価

48

ITT解析の解釈、問題点

49

- ▶ 治療遵守の程度を含めた、実践的医療技術の評価になるといわれる...
 - ▶ 臨床試験と実臨床は大きく異なる集団
 - ▶ Pragmatic 試験、市販後臨床試験実施が適切
- ▶ 非劣性試験では問題が生じることも
 - ▶ 全員が割付を無視して同じ治療を行えば、治療効果に差がない、ことを示せる
- ▶ ITT原則に従ったFAS集団とPPS集団の解析結果が大きく変わらないことを確認

49

アウトカムの比較結果

50

- 17a** 主要/副次評価項目の介入結果
 - ▶ 検定だけでなく、推定(信頼区間)もあわせて
- 17b** 二値アウトカムの場合、差と比の両方
 - ▶ 相対指標：リスク比、オッズ比
 - ▶ 絶対指標：リスク差

50

サブグループ解析結果

51

- 18** サブグループ解析、調整解析
 - ▶ Pre-specified か Post-hoc かを明らかに
 - ▶ Forest Plot

51

安全性

52

- 19** 有害事象と意図しない効果
 - ▶ 有害事象の報告が少ない
 - ▶ さらに試験治療による寄与も明らかでない
 - ▶ CONSORT拡張版で有害事象報告におけるチェックリストを提供
 - ▶ Ioannidis JP, et al. *Ann Intern Med* 2004.
 - ▶ 有効性と安全性両方の議論を行うべき

52

考察、その他

53

- 20** 限界
- 21** 一般化可能性
- 22** 解釈
- 23** 臨床試験登録番号と試験登録名
- 24** プロトコルの入手方法
 - ▶ 雑誌によってはプロトコルの提出は義務
- 25** 資金提供者、薬剤提供者
 - ▶ 資金提供者の役割も
 - ▶ 利益相反(Conflict of Interest)とはまた別

53

CONSORTに足りないもの

54

- ▶ IRBの承認
- ▶ 施設モニタリングがなされたか
- ▶ CONSORTはこれらを論文に盛り込むことを阻むものではない
 - ▶ 次回改訂で再度議論されるであろう

54

臨床試験論文を読む／書く上で

55

- ▶ CONSORTの説明論文を読む
 - ▶ BMJ 2010;340:c869
 - ▶ 文例まで載っており、論文を書く際に参考に
- ▶ 雑誌の特徴に注目
 - ▶ 商業誌か学会誌か、IF、論文の掲載時期
- ▶ それでも出版バイアスはなくなるらない
 - ▶ 結果の迅速な報告は義務となりつつある
 - ▶ 効きすぎた結果には疑いの目を

55

臨床試験を計画する上で

56

- ▶ 総括報告書や論文を念頭に置いたプロトコルの作成が要求される
 - ▶ 論文公表の際にSupplementとしてプロトコルと解析計画書の添付が要求される
 - ▶ しかも、1.0版とそこからの変更履歴も全て
- ▶ プロトコルのひな形
 - ▶ SPIRIT
 - ▶ JCOGのプロトコルテンプレート

56

Retire Statistical Significance

57

Amrhein V et al. Nature 2019;305-7.

57

NEJMの投稿規定が改訂

58

58

P値について

59

- ▶ 事前に定めた解析計画に従うことが重要
- ▶ 効果や関連を主張するために統計的な閾値（検定）を使うことは、type-Iエラーを制御するための方法を解析計画で述べた場合だけに限るべき
- ▶ 治療や曝露の有効性や害に関するエビデンスについて、点推定値とともに誤差の幅（信頼区間）も示すべき

59

まとめ

60

- ▶ 臨床試験登録
 - ▶ 出版バイアスを減らす取り組み
- ▶ 公正な出版のためのCONSORT
 - ▶ チェックリストとフローチャート
 - ▶ サンプルサイズ設計と検定
 - ▶ 解析対象集団、ITTの原則

60